

# istat working papers

N. 16  
2011

## **Il piano di selezione delle aziende agricole RICA nel campione RICA-REA 2010**

*Veronica Rondinelli*



# istat working papers

N. 16  
2011

## **Il piano di selezione delle aziende agricole RICA nel campione RICA-REA 2010**

*Veronica Rondinelli*

## **Comitato di redazione**

*Coordinatore:* Giulio Barcaroli

*Componenti:*

Rossana Balestrino	Francesca Di Palma	Luisa Picozzi
Marco Ballin	Alessandra Ferrara	Mauro Politi
Riccardo Carbini	Angela Ferruzza	Alessandra Righi
Claudio Ceccarelli	Danila Filipponi	Luca Salvati
Giuliana Coccia	Cristina Freguja	Giovanni Seri
Fabio Crescenzi	Aurea Micali	Leonello Tronti
Carla De Angelis	Nadia Mignolli	Sonia Vittozzi

*Segreteria:*

Lorella Appolloni, Maria Silvia Cardacino, Laura Peci, Gilda Sonetti, Antonio Trobia

## **Istat Working Papers**

Il piano di selezione delle aziende agricole RICA nel campione RICA-REA 2010

N. 16/2011

ISBN 88-458-1703-2

Istituto nazionale di statistica  
Servizio Editoria  
Via Cesare Balbo, 16 – Roma

## Il piano di selezione delle aziende agricole RICA nel campione RICA-REA 2010

Veronica Rondinelli<sup>1</sup>

### Sommario

*Il presente lavoro illustra la strategia campionaria utilizzata per la selezione delle aziende RICA del campione dell'indagine sui Risultati Economici delle Aziende agricole (RICAREA) 2010. Si tratta di una selezione che ha dato priorità alle aziende già selezionate per l'indagine dell'anno 2007. L'obiettivo è stato evitare di avere nella fase post-rilevazione un elevato tasso di mancata risposta. Da un punto di vista metodologico si è posto il problema di attribuire alle aziende incluse nel campione un riporto all'universo che tenesse in considerazione la diversa probabilità di inclusione. A tal scopo diverse sono state le sperimentazioni fatte. In questo lavoro particolare enfasi è data alla strategia di selezione delle aziende RICA e alla determinazione dei coefficienti di riporto all'universo delle unità selezionate.*

### Abstract

*This paper illustrates the sample strategy used to the selection of the RICA farms in the Farm Business Survey (RICA-REA) 2010. This is a selection that gets priority to the farms that selected in the survey 2007. The aim is to avoid having the post detection a high unit non-response rate. By a methodological point of view the problem is to attribute to the selected farms some weights that take into account the different inclusion probabilities. For this purpose several experiments are done. This work shows the experiment considered most appropriate for the problem: the selection of RICA farms that selected in 2007.*

**Parole chiave:** campione stratificato, legge delle probabilità composte, cut-off sampling, legge delle probabilità totali, probabilità di inclusione, Risultati Economici delle Aziende Agricole (RICA-REA)

### 1. Introduzione

A partire dall'annata agraria 2010 per l'indagine sui Risultati Economici delle Aziende agricole (RICA-REA)<sup>2</sup> è previsto l'utilizzo di un nuovo campione che tenga in considerazione le novità introdotte a livello comunitario in tema di agricoltura: la nuova tipologia di aziende agricole (Commissione Europea, 2008), la sostituzione della Unità di Dimensione Economica (UDE)<sup>3</sup> con un nuovo parametro: lo Standard Output<sup>4</sup> (European Commission European Commission, 2009). Proprio in virtù di questo cambiamento si è reso necessario ridisegnare il campione dell'indagine.

<sup>1</sup> Veronica Rondinelli Istat, e-mail: veronica.rondinelli@istat.it.

<sup>2</sup> Per ulteriori informazioni sull'indagine, si veda ISTAT (2002), Indagine sui Risultati Economici delle aziende agricole RICA-REA. Anni 2002-2003-2004. Istruzioni per la rilevazione.

<sup>3</sup> La UDE è l'Unità di Dimensione Economica di un'azienda agricola. Una UDE corrisponde a circa 1.200 € di Reddito Lordo Standard aziendale (RLS).

<sup>4</sup> All'art. 5 del Regolamento 1242/2008, lo standard output è definito "valore normale della produzione lorda". Esso è determinato per aree territoriali ed equivale alla somma dei valori ottenuti per ciascuna attività produttiva moltiplicando le produzioni standard unitarie per il numero di unità corrispondenti. L'incidenza percentuale della produzione standard delle diverse attività produttive dell'azienda rispetto alla sua produzione standard totale determina l'orientamento tecnico-economico aziendale (OTE). A seconda del livello di precisione dell'OTE si distinguono classi di OTE generali, principali e particolari.

In questo lavoro l'attenzione è focalizzata su una parte della strategia campionaria del nuovo campione RICA-REA 2010, ovvero la selezione delle aziende RICA.<sup>5</sup> Rispetto alle procedure equiprobabilistiche di selezione, il processo di selezione delle aziende RICA ha richiesto particolare attenzione. Si è adottata una selezione "coordinata" positivamente (Ballin et al., 2000), che ha dato priorità di selezione alle aziende che già avevano partecipato all'indagine nel 2007 (Rondinelli, 2007).<sup>6</sup>

Includere nel nuovo campione RICA-REA 2010 aziende agricole, già selezionate in passato, significa mantenere pressoché immutata la rete di rilevazione, con conseguenti effetti positivi in termini sia di qualità dei dati statistici raccolti che di riduzione del fastidio statistico.

La rete di rilevazione dell'indagine è gestita dall'INEA,<sup>7</sup> che raccoglie presso l'azienda informazioni economiche dettagliate attraverso il sistema di contabilità aziendale GAIA.<sup>8</sup> Non tutte le aziende agricole sono disponibili a fornire tali informazioni e, di conseguenza, è preferibile, proprio per motivi organizzativi e di qualità dei dati statistici raccolti, rilevare le aziende già presenti nella rete di rilevazione dell'INEA. Sostanzialmente proprio per evitare di registrare elevati tassi di mancata risposta è auspicabile che la maggior parte delle aziende selezionate abbiano partecipato alla rilevazione negli anni precedenti.

Altro motivo per selezionare aziende agricole che già hanno partecipato all'indagine negli anni precedenti, è la constatazione che a tutto oggi, purtroppo, non c'è un archivio aggiornato di aziende agricole. Il frame di riferimento per la selezione delle aziende RICA è, infatti, costituito dalle aziende del Censimento dell'Agricoltura 2000 con standard output maggiore uguale di 4.000 euro (un po' più di 985.000 aziende agricole).

Se selezionare le aziende agricole che già hanno partecipato all'indagine, è funzionale da un punto di vista della raccolta dati, non pochi sono i problemi di natura metodologica. Selezionare in primis le aziende agricole già presenti nell'indagine 2007 equivale ad escludere a priori dalla selezione quelle aziende che non vi hanno partecipato.

In letteratura l'esclusione di un set di unità dal frame di campionamento è nota come procedura di *cut-off sampling*.<sup>9</sup> In presenza di cut-off sampling si utilizza informazione ausiliaria per ridurre la potenziale distorsione delle stime. Solitamente si utilizzano tecniche di correzione dei pesi campionari con un coefficiente correttivo derivante dall'applicazione di un modello logistico (Haziza and al., 2009), di calibrazione (Deville and al., 1992) e di campioni bilanciati (Deville and al., 2004).

Nel caso in esame la selezione delle aziende RICA ha richiesto:

- l'attribuzione alle aziende selezionate di un peso campionario;
- la successiva *calibrazione* dei coefficienti di riporto all'universo.

Nel paragrafo a seguire ne sarà data illustrazione.

<sup>5</sup> Il campione RICA è una delle due parti in cui si compone il campione RICA-REA: la parte delle aziende small o REA (con meno di 4000 euro di standard output) e la parte, appunto, delle aziende commercial o RICA (con oltre 4000 euro di standard output). Il campione RICA è un campione stratificato, le cui variabili di stratificazione sono: la collocazione territoriale, lo standard output e l'orientamento tecnico economico (OTE). La strategia campionaria per la determinazione del campione RICA 2010 prevede la copertura di almeno il 90 per cento dello standard output complessivo delle aziende agricole. Con un apprezzabile livello di significatività statistica, ovvero con un errore campionario non superiore al 3 per cento a livello nazionale sulle variabili strategiche (costi, produzione, ecc.), è stata determinata, attraverso il processo di ottimizzazione dell'algoritmo genetico (GeneticStrata.R), una numerosità campionaria (14.103) tale da contenere i costi di rilevazione e, nel contempo, ridurre l'errore non campionario atteso delle stime e l'entità della molestia statistica.

<sup>6</sup> L'ultimo disegno di campionamento della RICA-REA adottato per gli anni contabili 2007, 2008 e 2009 è quello progettato nel 2007.

<sup>7</sup> L'INEA (Istituto di Economia Agraria) è titolare della parte RICA dell'indagine sui Risultati Economici delle Aziende Agricole (RICA-REA), mentre all'Istituto Nazionale di Statistica (ISTAT) della parte REA. L'ISTAT cura soprattutto gli aspetti metodologici, all'INEA spetta invece la cura della fase di raccolta dei dati.

<sup>8</sup> Per la raccolta dei dati presso le aziende agricole l'INEA utilizza il software contabile GAIA (Gestione Aziendale delle Imprese Agricole).

<sup>9</sup> La procedura di *cut-off sampling* consiste nell'esclusione di unità dell'universo dalla selezione campionaria. Le motivazioni possono essere molteplici: il loro contributo alle stime campionarie è minimo; il loro costo di rilevazione è molto alto. Nel caso in esame dell'indagine RICA-REA, si escludono dalla selezione delle aziende RICA quelle che non sono state selezionate per l'edizione precedente dell'indagine. La suddetta esclusione comporta di contenere le mancate risposte e di assicurare un numero consistente di aziende RICA, che l'INEA deve inviare alla DG-AGRI della Commissione Europea per la Rete di Informazione Contabile Europea.

## 2. La strategia di selezione delle aziende agricole RICA

Le diverse sperimentazioni fatte per la selezione delle aziende agricole RICA del campione 2010 hanno portato ad adottare una *selezione con sovrapposizione di campioni*.

La selezione casuale<sup>10</sup> delle aziende agricole è stata scartata in quanto appena 808 risultavano già essere state selezionate per l'indagine nel 2007 (Tavola 1).

**Tavola 1 - Campione delle aziende RICA selezionate per regione**

REGIONI / PROVINCE AUTONOME	Aziende	
	Estrate	Di cui selezionate per la RICA-REA 2007
Piemonte	769	39
Valle D'Aosta / Valleé D'Aoste	200	27
Lombardia	825	47
Veneto	940	54
Friuli Venezia Giulia	668	61
Liguria	659	54
Emilia Romagna	1.28	56
Toscana	799	60
Umbria	653	54
Marche	631	50
Lazio	742	28
Abruzzo	604	39
Molise	456	46
Campania	781	30
Puglia	896	17
Basilicata	571	41
Calabria	608	22
Sicilia	867	15
Sardegna	704	38
Trento	331	14
Bolzano/Bolzen	371	16
<b>ITALIA</b>	<b>14.103</b>	<b>808</b>

Si è deciso di utilizzare una selezione con sovrapposizione di campioni, includendo nel campione 2010 in primis le aziende che hanno partecipato all'indagine RICA-REA del 2007 e, solo in un secondo momento eventualmente, altre unità al fine di assicurare la numerosità teorica di strato prevista dal disegno di campionamento. Con la sovrapposizione dei campioni la casualità viene mantenuta essendo una caratteristica del primo piano di selezione adottato, che è mantenuto. Così facendo è risultato che quasi il 69 per cento delle aziende agricole del campione RICA sono già state selezionate nel campione del 2007 (Tavola 2).

Eseguita la selezione, il passo successivo è stato determinare il peso campionario per ciascuna unità estratta. Chiaramente si è esclusa l'ipotesi di attribuire all'interno di ciascuno strato, sia alle aziende estratte e selezionate all'indagine 2007 e sia alle aziende estratte casualmente, la stessa probabilità di inclusione:

$${}^h w = \frac{{}^h n}{{}^h N}$$

dove

${}^h n$  e  ${}^h N$  sono rispettivamente la numerosità di aziende agricole da campionare e la numerosità delle aziende RICA nell'universo limitatamente allo strato  $h$ -esimo ( $h=1,2,\dots,1001$ ) del campione RICA 2010.

<sup>10</sup> La selezione casuale è stata fatta attraverso l'utilizzo della generazione di numeri casuali.

Per la determinazione del coefficiente di riporto all'universo si sono fatte diverse applicazioni, in questo lavoro si riporta quella ritenuta più significativa:

- determinazione dei coefficienti di riporto all'universo differenziati: per le unità selezionate anche nell'indagine 2007 e per quelle selezionate casualmente;
- applicazione, successivamente, del metodo di *calibrazione o post-stratificazione*.

## 2.1 Il riporto all'universo

Per il riporto all'universo delle aziende agricole RICA, a ciascuna azienda si è attribuito un peso diversificato a secondo che l'azienda avesse già partecipato all'indagine nel 2007 oppure che fosse selezionata casualmente per la prima volta.

Selezionare in ogni strato *h-esimo* ( $h=1,2,\dots,1001$ ) del campione RICA 2010 le aziende che hanno partecipato all'indagine 2007 ( ${}^h n_{2007}$ ) comporta associare loro un peso campionario ( ${}^h w'_{iniziale\_2010}$ ) diverso da quello ( ${}^h w''_{iniziale\_2010}$ ) associato alle altre unità selezionate nello stesso strato *h-esimo* ma che sono estratte casualmente ( ${}^h n_{casuali}$ ), la cui estrazione è necessaria per coprire la numerosità teorica del campione ( ${}^h n$ ) nello strato *h-esimo*:  ${}^h n = {}^h n_{2007} + {}^h n_{casuali}$ .

Il peso campionario  ${}^h w''_{iniziale\_2010}$  delle unità estratte casualmente nello strato *h-esimo* è dato dal reciproco del rapporto tra la numerosità dello strato *h-esimo* ( ${}^h n$ ) e la differenza tra la numerosità dell'universo nello strato ( ${}^h N$ ) e la numerosità delle aziende selezionate per l'indagine del 2007 ( ${}^h N - {}^h n_{2007}$ ):

$${}^h w''_{iniziale\_2010} = \left( \frac{{}^h n}{{}^h N - {}^h n_{2007}} \right)^{-1}$$

Per le unità che hanno già partecipato all'indagine 2007, di numerosità  ${}^h n_{2007}$  nello strato *h-esimo*, il calcolo del loro peso campionario  ${}^h w'_{iniziale\_2010}$  è pari all'inverso della rispettiva probabilità di inclusione ( ${}^h \varphi$ ) delle suddette unità:

$${}^h w'_{iniziale\_2010} = ({}^h \varphi)^{-1}$$

Per il riporto all'universo delle aziende selezionate con sovrapposizione dei campioni si sono sfruttate le proprietà degli eventi e le leggi della probabilità: Legge delle Probabilità Totali e Legge delle Probabilità Composte.

Sia *A* l'evento:

*A* = (azienda *i*-esima dello strato *h-esimo* selezionata nel campione RICA 2010 e selezionata per la RICA-REA 2007)  $(i=1, \dots, 9.728, h=1,2,\dots,1001)$

La probabilità di questo evento (Dall'Aglio, 1995) può essere calcolata utilizzando la seguente relazione:

$$\text{Prob}(A) = \text{Pr ob}(A|I) * \text{Pr ob}(I) + \text{Pr ob}(A|\bar{I}) * \text{Pr ob}(\bar{I})$$



dove

$I =$  (l'azienda  $i$ -esima è stata selezionata per la RICA-REA 2007) e

$\bar{I} =$  (l'azienda  $i$ -esima non è stata selezionata per la RICA-REA 2007 e relativamente allo strato  $h$ -esimo si ha che:

$$\Pr ob(A|I) = 1$$

$$\Pr ob(I) = \left( {}^h w_{campione\_2007} \right)^{-1}$$

$$\Pr ob(A|\bar{I}) = \frac{{}^h n - {}^h n_{2007}}{{}^h N - {}^h n_{2007}}$$

$$\Pr ob(\bar{I}) = \left\{ 1 - \left( {}^h w_{campione\_2007} \right)^{-1} \right\}$$

Quindi nello strato  $h$ -esimo la probabilità di inclusione nel campione ( ${}^h \varphi$ ) per ogni unità  $i$ -esima selezionata per l'indagine 2007 è:

$$({}^h \varphi) = 1 * \left( {}^h w_{campione\_2007} \right)^{-1} + \left( \frac{{}^h n - {}^h n_{2007}}{{}^h N - {}^h n_{2007}} \right) * \left[ 1 - \left( {}^h w_{campione\_2007} \right)^{-1} \right].$$

La presenza di variabilità nella stima del numero di aziende dell'universo è corretta attraverso un processo di post-stratificazione:

per ogni strato  $h$ -esimo si calcola il fattore correttivo post - stratificazione  $\frac{{}^h N}{\sum {}^h w_{iniziale\_2010}}$  che

va a moltiplicare il corrispettivo peso campionario.

In ogni strato  $h$ -esimo, quindi, il peso iniziale delle aziende RICA 2010 ivi appartenenti è dato dal prodotto del peso iniziale per il fattore correttivo post-stratificazione:

$${}^h w_{-2010} = {}^h w_{iniziale\_2010} * \frac{{}^h N_{-2010}}{\sum {}^h w_{iniziale\_2010}}$$

La somma dei suddetti pesi stima l'universo di riferimento delle aziende RICA 2010 (Tavola 2).

**Tavola 2 - Campione RICA 2010 e universo stimato e post-stratificazione a livello regionale**

REGIONI / PROVINCE AUTONOME	Aziende		Universo	
	Totale	Di cui selezionate per la RICA-REA 2007	Stima del numero di aziende(*)	Stima del numero di aziende post- stratificazione
Piemonte	769	544	41.224	61.109
Valle D'Aosta / Valleé D'Aoste	200	154	1.289	2.044
Lombardia	825	559	24.638	47.426
Veneto	940	672	47.241	71.919
Friuli Venezia Giulia	668	440	4.666	13.893
Liguria	659	358	9.444	14.603
Emilia Romagna	1.028	716	35.973	71.474
Toscana	799	620	29.572	43.160
Umbria	653	488	10.297	15.563
Marche	631	528	10.555	27.046
Lazio	742	479	34.052	58.214
Abruzzo	604	446	21.587	37.432
Molise	456	349	7.325	14.240
Campania	781	543	65.238	92.953
Puglia	896	502	107.311	120.133
Basilicata	571	405	12.887	22.242
Calabria	608	396	56.597	74.093
Sicilia	867	524	101.713	123.878
Sardegna	704	475	31.757	42.356
Trento	331	263	10.542	18.364
Bolzano/Bolzen	371	267	9.152	13.435
<b>ITALIA</b>	<b>14.103</b>	<b>9.728</b>	<b>673.061</b>	<b>985.577</b>

(\*) La stima del numero di aziende è data dalla somma dei pesi iniziali.

### Notazioni conclusive

Alla luce di quanto fin qui esposto, si possono trarre alcune considerazioni in merito alla selezione delle aziende RICA 2010. Questo lavoro rappresenta un punto di soluzione a un problema abbastanza complesso: da una parte le esigenze pratiche della rilevazione (mantenere pressoché immutata la rete di rilevazione, evitare di avere molte mancate risposte, ridurre il disturbo statistico, ecc.) e, dall'altra le esigenze metodologiche (affidabilità delle stime prodotte per gli aggregati economici oggetto di studio). La procedura, ritenuta più adatta per la soluzione del problema, è stata la selezione delle aziende RICA, che può essere vista come un esempio di applicazione di *cut-off sampling* al frame di campionamento. In ciascuno strato si sono incluse con certezza nel campione le aziende già selezionate per la RICA-REA 2007 e, in un secondo momento, si sono estratte casualmente le restanti aziende necessarie per coprire la numerosità campionaria di strato pianificata nel disegno di campionamento. I coefficienti di riporto all'universo sono stati determinati in modo differenziato. Per le aziende RICA selezionate e già presenti nell'indagine RICA-REA 2007 i pesi sono stati determinati attraverso l'applicazione delle leggi probabilistiche; per le rimanenti aziende RICA, selezionate casualmente, il peso campionario è stato, invece, determinato nello strato come inverso della rispettiva probabilità di inclusione. I coefficienti di riporto all'universo, così calcolati, hanno poi successivamente subito un processo di *calibrazione* per la evitare il problema di sottostima.

## Riferimenti bibliografici

- Ballin, M. e Loriga, S. (2000), *Metodi di selezione coordinata nelle indagini campionarie sulle imprese*, Rivista di Statistica Ufficiale, Fascicolo 2, Franco Angeli.
- Commissione Europea (2008), *REGOLAMENTO (CE) N. 1242/2008 DELLA COMMISSIONE dell'8 dicembre 2008 che istituisce una tipologia comunitaria delle aziende agricole*, Brussels.
- Dall'Aglio, G. (1995), *Calcolo delle Probabilità*, Zanichelli Editore, Bologna.
- Deville, J.C. and Särndal, C. E. (1992), *Calibration Estimation in Survey Sampling*, Journal of the American Statistical Association, Vol. 87, No. 418.
- Deville, J-C. and Tillé, Y. (2004), *Efficient balanced sampling: the Cube method*, Biometrika, 91.
- European Commission, Directorate General for Agriculture and rural development (2009), *Community Committee for The Farm Accountancy Data Network, RI/CC 1500 rev. 3*, Brussels, 05/10/2009.
- ISTAT (2002), *Indagine sui Risultati Economici delle aziende agricole RICA-REA. Anni 2002-2003-2004. Istruzioni per la rilevazione*.
- ISTAT (2005), Genesees v. 3.0. "Funzione Riponderazione". Manuale utente e aspetti metodologici.
- Rondinelli, V. (2007), *Nota Tecnica: Disegno di campionamento dell'indagine RICA-REA. Anno 2007*, Documento interno ISTAT.
- Van Huis M., Koeijers E., De Ree J. (1994), *Response Burden and Coordinated Sampling for Economic Surveys*, Netherlands Official Statistics, Vol. 9, pp. 17-26.

## Informazioni per gli autori

La collana è aperta ad autori dell'Istat e del Sistema statistico nazionale, e ad altri studiosi che abbiano partecipato ad attività promosse dal Sistan (convegni, seminari, gruppi di lavoro, ecc.). Da gennaio 2011 essa sostituirà Documenti Istat e Contributi Istat.

Coloro che desiderano pubblicare sulla nuova collana dovranno sottoporre il proprio contributo alla redazione degli Istat Working Papers inviandolo per posta elettronica all'indirizzo [iwp@istat.it](mailto:iwp@istat.it). Il saggio deve essere redatto seguendo gli standard editoriali previsti, corredato di un sommario in italiano e in inglese; deve, altresì, essere accompagnato da una dichiarazione di paternità dell'opera. Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del *Chicago Manual of Style*.

Per gli autori Istat, la sottomissione dei lavori deve essere accompagnata da una mail del proprio dirigente di Servizio/Struttura, che ne assicura la presa visione. Per gli autori degli altri enti del Sistan la trasmissione avviene attraverso il responsabile dell'ufficio di statistica, che ne prende visione. Per tutti gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione. Tutti i lavori saranno sottoposti al Comitato di redazione, che valuterà la significatività del lavoro per il progresso dell'attività statistica istituzionale. La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line.

Gli articoli pubblicati impegnano esclusivamente gli autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat. Si autorizza la riproduzione a fini non commerciali e con citazione della fonte.